

# ブートストラップ推定法による 相関係数の信頼区間（非正規な場合）

Bootstrap Confidence intervals for the correlation coefficient in non-normal cases

松 田 忠 之  
Matsuda, Tadayuki

## ABSTRACT

Recent advances in bootstrap methods allow the construction of highly accurate approximate confidence intervals, even for very complicated probability models. The present paper is devoted to an experimental study of confidence intervals for the correlation coefficient of samples from non-normal populations. We shall investigate the robustness of confidence intervals for the correlation coefficient based on four bootstrap procedures: bootstrap-t, percentile, BCa, and ABC when populations vary widely in the degree of non-normality. Also we show that the calibration technique can not improve the coverage accuracy of bootstrap confidence intervals in these non-normal cases.

## 1 序

表 1 は「医学への統計学」（丹後[7], 表 10）より引用した, 76 名の人工透析患者のヘモグロビン HB と赤血球数 RBC のデータである。HB と RBC の相関係数は約 0.8186 である。これらのデータが 2 次元正規分布にしたがうと仮定できるならば, 相関係数の信頼係数 95% の信頼区間の理論値は (0.7273, 0.8814) である。ところが, ブートストラップ法を使って求めた相関係数の信頼区間は理論値と少し異なる（松田[8]）。この原因は, おそらく HB と RBC の同時分布が正規分布にしたがわないことによると考えられる。

表1 人工透析患者のヘモグロビン HB (mg/dl) と赤血球数 RBC ( $\times 10^4$ )

HB	RBC	HB	RBC	HB	RBC	HB	RBC	HB	RBC
6.2	204	5.6	187	4.8	140	9.0	273	7.7	226
6.5	221	5.3	174	6.2	176	9.0	294	5.3	140
8.8	289	4.6	156	5.5	170	5.2	158	6.8	177
9.6	315	5.9	173	7.6	235	6.4	186	7.0	270
7.0	239	6.1	237	5.7	224	5.5	161	11.3	371
6.4	218	7.4	197	4.9	223	5.7	175	5.8	280
7.8	247	6.8	192	6.8	273	9.2	286	6.5	287
5.9	176	6.9	208	7.8	234	9.6	313	6.2	196
11.8	341	8.1	260	9.4	314	10.0	285	7.0	233
6.1	190	6.9	242	3.9	121	7.6	227	6.2	193
5.6	158	6.8	242	6.2	211	8.0	286	5.2	165
6.4	239	7.6	208	6.1	197	8.3	290	8.4	309
6.6	213	6.0	200	7.2	104	7.6	183	8.7	297
8.0	234	6.4	167	5.7	175	7.8	239	7.2	265
8.9	257	6.6	204	7.7	248	5.9	139	7.6	260
6.3	187								

そこで、データを正規確率紙にプロットして周辺分布の正規性について調べてみる。図1と図2は、それぞれHBとRBCのデータをGNUの統計ソフトRを使って正規確率紙にプロットしたものである。データ(○印)が図中の直線に近いときは、正規性が成り立つと考えられる。図1から、HBは値の小さいところと大きいところで直線からの乖離が大きく、一方図2から、RBCは値の小さいところで直線から外れているのがわかる。

さらに、それぞれの正規性を調べるために、統計解析プログラムパッケージSASを使って仮説検定を行う。表2と表3はHBとRBCについて正規性の検定を行った結果である。

HBの正規性検定では、シャピロ・ウィルク検定統計量の値は $W = 0.957643$ 、またその $p$ 値( $Pr < W$ )は0.0125である。したがって、仮説「母集団は正規分布である」は有意水準5%で棄却される。同様に、コルモゴロフ・スミルノフ検定統計量の値は $D = 0.105613$ 、その $p$ 値( $Pr > D$ )は0.0354となり、やはり仮説は棄却される。一方、RBCの正規性検定では、いずれの検定によっても仮説が採択される。

上記の検定結果から、HBとRBCの2次元データの正規性は成立しない。こ

図1 HB の正規確率プロット

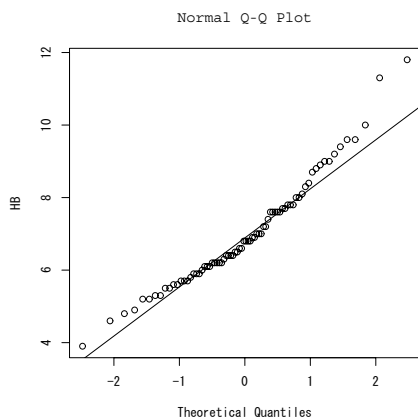


図2 RBC の正規確率プロット

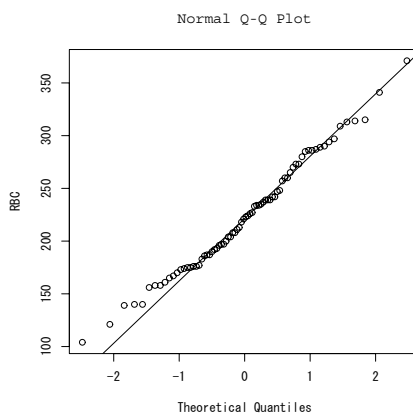


表2 HB の正規性検定

Test	Staistic	p Value
Shapiro-Wilk	W 0.957643	Pr<W 0.0125
Kolmogorov-Smirnov	D 0.105613	Pr>D 0.0354

表3 RBC の正規性検定

Test	Staistic	p Value
Shapiro-Wilk	W 0.986634	Pr<W 0.6103
Kolmogorov-Smirnov	D 0.067955	Pr>D >0.1500

の例のように母集団分布が正規分布でない場合に相関係数の信頼区間を求める方法として、ブートストラップ- $t$ 法、パーセンタイル法、BCa法、ABC法などのノンパラメトリック・ブートストラップ法（bootstrap method）の理論的研究が近年盛んに行われている（Efron[4]、Efron and Tibshirani[5]、Davison and Hinkley[1]）。

本稿では、母集団分布が正規分布に近い分布からまったく異なる分布まで連

続的に変化するとき、シミュレーション実験によってデータを人工的に生成し、上述のノンパラメトリック・ブートストラップ法で推定した相関係数の信頼区間の頑健性について調べる。さらに、ブートストラップ法による推定値を補正する calibration がどの程度有効に働くかについても調べる。ここで考察する母集団は、正規分布、および、正規分布とガンマ分布の混合分布を周辺分布としてもつ分布である。この分布は、ガンマ分布の shape パラメータを変化させることによって、正規分布に非常に近い分布からまったく異なる分布まで連続的に動く。

## 2 正規分布、および、正規分布とガンマ分布の混合分布の相関係数の信頼区間

確率変数  $U$  の密度関数が

$$p(u) = \frac{(u-\gamma)^{\alpha-1} \exp\left\{-\frac{u-\gamma}{\beta}\right\}}{\beta^{\alpha} \Gamma(\alpha)}, \quad \alpha > 0, \beta > 0, u > \gamma$$

で与えられるガンマ分布を一般に  $\Gamma(\alpha, \beta, \gamma)$  で表す。 $\alpha$  はガンマ分布の shape パラメータと呼ばれる。この論文では、 $U$  はガンマ分布  $\Gamma(\alpha, \alpha^{-1/2}, -\alpha^{1/2})$  にしたがうと仮定する。このとき、 $U$  の平均、分散、歪度 (skewness)、尖度 (kurtosis) は、それぞれ  $0, 1, 2\alpha^{-1/2}, 3+6\alpha^{-1}$  で与えられる。したがって、 $\alpha$  の値が大きくなれば、 $U$  は標準正規分布に収束する。しかし、 $\alpha$  の値が小さいとき、正規分布とはまったく異なる分布である。

いま、 $Z$  を  $U$  と独立な標準正規分布にしたがう確率変数とする。このとき、 $0 < \rho < 1$  に対して、2 次行列を

$$C = \begin{pmatrix} 1 & 0 \\ \rho & \sqrt{1-\rho^2} \end{pmatrix}$$

によって定義し、2 次元確率変数  $(Z, U)'$  の  $C$  による変換を

$$X = C \begin{pmatrix} Z \\ U \end{pmatrix} = \begin{pmatrix} Z \\ \rho Z + \sqrt{1-\rho^2} U \end{pmatrix}$$

で表す。変換された確率変数  $X$  の第 1 成分  $X_1$  は標準正規分布  $Z$  に等しく、第 2 成分  $X_2$  は標準正規分布とガンマ分布の混合分布である。このとき、 $X = (X_1, X_2)'$  の平均ベクトルと共分散行列はそれぞれ

$$E(X) = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \text{Cov}(X) = \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}$$

で与えられ、とくに  $X_1$  と  $X_2$  の相関係数は  $\rho$  である。

確率変数  $X_1$  と  $X_2$  の相関係数  $\rho$  の信頼係数 95% 信頼区間をモンテカルロ法を使ったシミュレーション実験で求めよう。この実験において生成するデータは、表 1 のデータと同じ標本の大きさ  $n = 76$  と相関係数  $\rho = 0.82$  をもつと仮定する。ガンマ分布の shape パラメータ  $\alpha$  の値を 0.1, 0.5, 1, 4, 9, 25 の 6 通りの場合に、2 次元確率変数  $X$  の分布にしたがう 76 個の乱数を独立に生成して標本相関係数  $r$  を求め、この操作を 100,000 回繰り返して、 $\rho$  の信頼係数 95% の信頼区間を求めた。参考のために、これら 100,000 個の相関係数のヒストグラムを付図 1 に示した。図中の alpha は  $\alpha$  の意味で使用した。

表 4 には、信頼区間の 2.5% 点と 97.5% 点をそれぞれ 0.025 の欄と 0.975 の欄に示した。これらの値は信頼区間の理論値と考えられ、次節のブートストラップ法で推定した信頼区間の精度を調べるために利用できる。さらに表には、 $r$  の平均と標準偏差を計算して示した。

ガンマ分布の shape パラメータ  $\alpha$  の値が  $\infty$  のとき、確率変数  $X$  は 2 次元正規

表 4 相関係数の信頼係数 95% の信頼区間

$\alpha$	0.025	0.975	$r$ の平均	$r$ の標準偏差
0.1	0.58681	0.97055	0.83641	0.10095
0.5	0.67700	0.91927	0.82288	0.06265
1	0.69991	0.90430	0.82062	0.05272
4	0.72163	0.88921	0.81881	0.04289
9	0.72725	0.88608	0.81824	0.04069
25	0.73066	0.88427	0.81853	0.03937
$\infty$	0.72939	0.88234		

分布を表す。表から2次元正規分布の信頼区間の理論値は $\alpha = 25$ の場合の値とそれほど違わない。ところが、 $\alpha$ の値が小さくなり $X$ の正規性が崩れると、信頼区間の下限値が小さくなり信頼区間の幅は大きくなる。

相関係数 $\rho$ の推定値 $r$ は偏りを持つことが知られている。Gayen[6]によれば、 $r$ の平均と分散の近似式は

$$E(r) \doteq \rho + \frac{1}{n} \left\{ -\frac{1}{2}\rho(1-\rho^2) + \frac{1}{8}L_{4,1} \right\}$$

$$V(r) \doteq \frac{1}{n} \left\{ (1-\rho^2)^2 + \frac{1}{4}L_{4,2} \right\}$$

で与えられる。ここで、 $L_{4,1}$ ,  $L_{4,2}$ は、2次元分布のキュミュラント $\kappa_{ij}$ の比率

$$\gamma_{ij} = \frac{\kappa_{ij}}{\sqrt{\kappa_{20}^i \kappa_{02}^j}}$$

と $\rho$ の関数として次のように定義される。

$$L_{4,1} = 3\rho(\gamma_{40} + \gamma_{04}) - 4(\gamma_{31} + \gamma_{13}) + 2\rho\gamma_{22} = 18\rho(1-\rho^2)^2/\alpha$$

$$L_{4,2} = \rho^2(\gamma_{40} + \gamma_{04}) - 4\rho(\gamma_{31} + \gamma_{13}) + 2(2+\rho^2)\gamma_{22} = 6\rho^2(1-\rho^2)^2/\alpha.$$

$\rho = 0.82$ ,  $n = 76$  とするとき、その平均と分散は

$$E(r) \doteq 0.818233 + 0.002605/\alpha$$

$$V(r) \doteq 0.001412 + 0.001424/\alpha$$

となり、個々の $\alpha$ に対する平均、分散および標準偏差は表5のとおりである。 $r$ の平均と標準偏差を表4のシミュレーション結果と比較すると、 $\alpha = 0.1$ の場合はそれらの値が若干異なるが、それ以外の $\alpha$ においてはほぼ一致していることがわかる。

表5  $r$ の平均、分散、標準偏差

$\alpha$	平均	分散	標準偏差
0.1	0.84429	0.01565	0.12512
0.5	0.82344	0.00426	0.06527
1	0.82084	0.00284	0.05326
4	0.81888	0.00177	0.04205
9	0.81852	0.00157	0.03963
25	0.81834	0.00147	0.03833

上の表から、 $\alpha = 0.1, 0.5$  のように小さい値のとき、 $r$  の平均は真の値 0.82 より大きく、逆に、 $\alpha$  の値が大きくなると、真の値より少し小さくなる。一方、 $r$  の分散は  $\alpha$  の値が大きくなるとしだいに小さくなる傾向がある。この偏りと分散がブートストラップ法によりどの程度推定可能か試してみよう。

まず、モーメント法を使ったブートストラップ法による偏りと分散の推定を紹介する。標本  $x = (x_1, \dots, x_n)$  から大きさ  $n$  の無作為標本を復元抽出することを  $B$  回反復し、 $B$  個のブートストラップ標本  $x_1^*, \dots, x_B^*$  からそれぞれの相関係数  $r_1^*, \dots, r_B^*$  を計算する。このとき、偏りは 1 次のモーメント

$$B_r = \frac{1}{B} \sum_{i=1}^B r_i^* - r = \bar{r}^* - r$$

によって推定され、また、分散は 2 次のモーメント

$$V_r = \frac{1}{B-1} \sum_{i=1}^B (r_i^* - \bar{r}^*)^2$$

によって推定される。

つぎに、ジャックナイフ推定法による偏りと分散の推定を紹介する。まず、与えられた標本  $x = (x_1, \dots, x_n)$  から  $i$  番目のデータを削除した  $(n-1)$  個のデータを

$$x_{(i)} = (x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n)$$

で表す。 $x_{(i)}$  から計算した相関係数  $\hat{r}_{(i)}$  を用いて、偏りと分散のジャックナイフ推定量をそれぞれ

$$\hat{bias}_{jack} = (n-1)(\hat{r}_{(.)} - r), \quad \hat{r}_{(.)} = \frac{1}{n} \sum_{i=1}^n \hat{r}_{(i)}$$

$$\hat{var}_{jack} = \frac{n-1}{n} \sum_{i=1}^n (\hat{r}_{(i)} - \hat{r}_{(.)})^2$$

によって定義する。

分散の 3 番目の推定方法として、影響関数 (influence function) を用いたノンパラメトリックデルタ法 (delta-method) を紹介する。まず、2 次元確率変数

$X = (X_1, X_2)'$  の  $X_1$  の  $r$  次,  $X_2$  の  $s$  次モーメントを  $\mu_{rs} = E(X_1^r X_2^s)$  で表せば,  $X_1, X_2$  の相関係数は

$$\rho = \frac{\mu_{11} - \mu_{10}\mu_{01}}{\sqrt{(\mu_{20} - \mu_{10}^2)(\mu_{02} - \mu_{01}^2)}}$$

と書ける。Davison and Hinkley[1] によれば,  $\rho$  の影響関数  $L_\rho(x)$  は

$$L_\rho(x) = \tilde{x}_1 \tilde{x}_2 - \frac{1}{2} \rho (\tilde{x}_1^2 + \tilde{x}_2^2), \quad x = (x_1, x_2)$$

である。ここで

$$\tilde{x}_1 = \frac{x_1 - \mu_{10}}{\sqrt{\mu_{20} - \mu_{10}^2}}, \quad \tilde{x}_2 = \frac{x_2 - \mu_{01}}{\sqrt{\mu_{02} - \mu_{01}^2}}.$$

標本  $x = (x_1, \dots, x_n)$  とその相関係数  $r$  が与えられたとき, ノンパラメトリックデルタ法による分散の推定値  $v_L$  は次の式を使って計算できる。

$$v_L = n^{-2} \sum_{i=1}^n \{ \tilde{x}_{i1} \tilde{x}_{i2} - \frac{1}{2} r (\tilde{x}_{i1}^2 + \tilde{x}_{i2}^2) \}^2$$

ただし,  $x_i = (x_{i1}, x_{i2})$ ,  $i = 1, \dots, n$ , に対して

$$\tilde{x}_{i1} = \frac{x_{i1} - \bar{x}_1}{\sqrt{\sum (x_{i1} - \bar{x}_1)^2 / n}}, \quad i = 1, \dots, n$$

$$\tilde{x}_{i2} = \frac{x_{i2} - \bar{x}_2}{\sqrt{\sum (x_{i2} - \bar{x}_2)^2 / n}}, \quad i = 1, \dots, n.$$

偏りと分散のブートストラップ推定法の比較を行うために, それぞれのブートストラップ法を 100 回繰り返し行なった。表 6 は, 100 回の偏りの推定値の平均とその標準偏差を計算した。なお, 左端の理論値は, 表 5 の平均から 0.82 を引いた値である。

同様に, 表 7 には, 各ブートストラップ法による 100 回の分散の平方根 (標準偏差) の平均値とそれらの標準偏差を示した。比較を容易に行うために表 5 の標準偏差の値を左端の理論値として書き加えた。

表 6 から,  $\alpha = 0.1, 0.5$  のとき, 偏りの推定はモーメント法の方がジャックナ



表 6 偏りの推定

推定法		ジャックナイフ法		モーメント法 $B_r$	
$\alpha$	理論値	平均	標準偏差	平均	標準偏差
0.1	0.02429	0.03188	0.00370	0.02169	0.00398
0.5	0.00344	0.00417	0.00151	0.00316	0.00237
1	0.00084	0.00088	0.00117	0.00065	0.00181
4	-0.00112	-0.00132	0.00083	-0.00134	0.00153
9	-0.00148	-0.00176	0.00078	-0.00170	0.00134
25	-0.00166	-0.00190	0.00082	-0.00201	0.00142

表 7 標準偏差の推定

推定法		デルタ法		ジャックナイフ法		モーメント法 $V_r$	
$\alpha$	理論値	平均	標準偏差	平均	標準偏差	平均	標準偏差
0.1	0.12512	0.11374	0.00439	0.15038	0.00757	0.10993	0.00480
0.5	0.06527	0.06325	0.00399	0.07040	0.00442	0.06261	0.00439
1	0.05326	0.05125	0.00362	0.05522	0.00358	0.05143	0.00403
4	0.04205	0.04083	0.00312	0.04286	0.00320	0.04202	0.00335
9	0.03963	0.03921	0.00295	0.04109	0.00309	0.04041	0.00337
25	0.03833	0.03751	0.00277	0.03921	0.00304	0.03874	0.00300

イフ法より若干良いが、それ以外の値ではあまり変わらない。しかし、標準偏差はモーメント法の方がジャックナイフ法より各  $\alpha$  について一様に少し大きい。さらに、表 7 から標準偏差の推定値としてはデルタ法が良く、たとえ  $\alpha$  の値が小さくなって正規性が崩れても安定した良い推定値を与えることがわかる。

### 3 ブートストラップ法による相関係数の信頼区間

相関係数  $\rho$  の信頼係数  $100(1-2a)\%$ ,  $0 < a < 1$ , パーセント信頼区間を求めるために、よく知られたブートストラップ法のアルゴリズムを紹介する。以下のアルゴリズムでは、標本  $x = (x_1, \dots, x_n)$  がすでに観測されており、 $x$  による  $\rho$  の推定値  $r$ , および  $r$  の標準偏差  $s$  を利用できると仮定する。

まず、相関係数をステューデント化してブートストラップ法を適用するブートストラップ- $t$  法のアルゴリズムを紹介する。

1. 標本  $x = (x_1, \dots, x_n)$  から大きさ  $n$  の無作為標本を復元抽出することを  $B$  回反復し、 $B$  個のブートストラップ標本  $x_1^* = (x_{11}^*, \dots, x_{1n}^*), \dots, x_B^* =$

- $(x_{B1}^*, \dots, x_{Bn}^*)$  を抽出する。
2. 各々のブートストラップ標本  $x_b^*(b=1, \dots, B)$  に対して,  $\rho$  の推定値  $\hat{r}_b^*$  とその標準偏差  $\hat{s}_b^*$  を推定し, スチューデント化された統計量  $\hat{t}_b^* = (\hat{r}_b^* - r) / \hat{s}_b^*$  を計算する。
  3.  $\hat{t}_b^*(b=1, \dots, B)$  の  $100a$  パーセント点  $\hat{t}^*(a)$  と  $100(1-a)$  パーセント点  $\hat{t}^*(1-a)$  を求める。
  4. このとき,  $\rho$  に対する信頼係数  $100(1-2a)$  パーセントの信頼区間は  $(r - \hat{t}^*(1-a)\hat{s}, r - \hat{t}^*(a)\hat{s})$  である。

なお, オリジナル標本やブートストラップ標本から  $r$  の標準偏差を推定する方法として, 前節で紹介したノンパラメトリックデルタ法, ジャックナイフ法, モーメント法などが利用できる。

つぎに, パーセンタイル法を直接使って相関係数  $\rho$  の信頼区間を求める方法を紹介する。

1. 標本  $x = (x_1, \dots, x_n)$  から大きさ  $n$  の無作為標本を復元抽出することを  $B$  回反復し,  $B$  個のブートストラップ標本  $x_1^* = (x_{11}^*, \dots, x_{1n}^*), \dots, x_B^* = (x_{B1}^*, \dots, x_{Bn}^*)$  を抽出する。
2. 各々のブートストラップ標本  $x_b^*(b=1, \dots, B)$  に対して, 相関係数  $\hat{r}_b^*$  を計算する。
3.  $\hat{r}_b^*(b=1, \dots, B)$  の  $100a$  パーセント点  $\hat{r}^*(a)$  と  $100(1-a)$  パーセント点  $\hat{r}^*(1-a)$  を求める。
4. このとき, 信頼係数  $100(1-2a)$  パーセントの信頼区間は  $(\hat{r}^*(a), \hat{r}^*(1-a))$  である。

標本相関係数  $r$  の分布は母集団の相関係数が 0 以外の場合は非対称になり, 正規分布に対する近似がよくない。ところが, フィッシャー (Fisher) の  $z$  変換

$$\hat{\phi} = \frac{1}{2} \log \frac{1+r}{1-r}$$

を行えば, データが 2 次元正規分布と考えられるとき,  $\hat{\phi}$  は平均  $\phi (= 1/2 \log$

$((1+\rho)/(1-\rho))$ , 分散  $1/(n-3)$  の正規分布によって近似できることが知られている。そこで, 変換した  $\hat{\phi}$  に対してブートストラップ- $t$  法を使って信頼区間を求め, つぎにその逆変換

$$\hat{\rho} = \frac{e^{2\hat{\phi}} - 1}{e^{2\hat{\phi}} + 1}$$

によって  $\rho$  の信頼区間を求めることができる。この方法は直接  $\rho$  の信頼区間を求めるより精度がよくなることが知られている。Davison and Hinkley [1] によれば,  $\phi = 1/2 \log((1+\rho)/(1-\rho))$  に対する影響関数は

$$L_{\phi}(x) = \frac{1}{1-\rho^2} \{ \bar{x}_1 \bar{x}_2 - \frac{1}{2} \rho (\bar{x}_1^2 + \bar{x}_2^2) \}, \quad x = (x_1, x_2)$$

で与えられ, ノンパラメトリックデルタ法ではこの影響関数を使って標準偏差の推定を行う。

なお,  $\rho$  に対するフィッシャーの  $z$  変換を使って,  $\phi$  に対するパーセンタイル法から信頼区間を求め, 逆変換で  $\rho$  に対する信頼区間を求める方法は,  $\phi$  に対してパーセンタイル法を直接適用する方法と同じ結果を導く。すなわち, パーセンタイル法は変換を考慮する必要がないという利点を持つ。

パーセンタイル法の改良として, BCa (bias-corrected and accelated) と呼ばれる方法を紹介する (Efron and Tibshirani [5])。

まず, 偏りを推定するために, パーセンタイル法のアルゴリズムで紹介したブートストラップ標本  $x_b^*$  から計算した相関係数  $\hat{r}_b^*$  を使って

$$\hat{z}_0 = \Phi^{-1} \left( \frac{\#\{\hat{r}_b^* < r\}}{B} \right)$$

を定義する。ここで,  $\Phi$  は標準正規分布の分布関数を表し,  $\Phi^{-1}$  はその逆関数である。 $\hat{z}_0$  は偏り修正 (bias-correction) と呼ばれ,  $r$  より小さい  $\hat{r}_b^*$  の個数が丁度半分 ( $B/2$  個) のとき 0 となる。

相関係数  $\hat{r}_b^*$  ( $b = 1, \dots, B$ ) の  $100a$  パーセント点を  $\hat{r}^*(a)$  とする。このとき, BCa の信頼係数  $100(1-2a)$  パーセントの信頼区間は

$$(\hat{r}^*(a_1), \hat{r}^*(a_2))$$

である。ただし、 $a_1$  と  $a_2$  は次の関係式を使って求める。

$$a_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z(a)}{1 - \hat{a}(\hat{z}_0 + z(a))}\right)$$

$$a_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z(1-a)}{1 - \hat{a}(\hat{z}_0 + z(1-a))}\right)$$

ここで  $z(a)$  と  $z(1-a)$  はそれぞれ標準正規分布の  $100a$  パーセント点と  $100(1-a)$  パーセント点を表す。とくに、 $\hat{z}_0 = 0$ ,  $\hat{a} = 0$  と仮定すれば、

$$a_1 = \Phi(z(a)) = a$$

$$a_2 = \Phi(z(1-a)) = 1-a$$

を満足し、BCa 法による信頼区間はパーセンタイル法による信頼区間と等しくなることがわかる。

$\hat{a}$  は acceleration と呼ばれる量であり、歪度 (skewness) の  $1/6$  倍を表し、その推定には影響関数やジャックナイフ法が使われる。影響関数を使った推定式は次式で与えられる。

$$\hat{a} = \frac{1}{6} \frac{\sum_{i=1}^n L_\rho(x_i)^3}{\{\sum_{i=1}^n L_\rho(x_i)^2\}^{3/2}}$$

一方、ジャックナイフ法による推定は

$$\hat{a} = \frac{1}{6} \frac{\sum_{i=1}^n (\hat{r}_{(.)} - \hat{r}_{(i)})^3}{\{\sum_{i=1}^n (\hat{r}_{(.)} - \hat{r}_{(i)})^2\}^{3/2}}$$

によって与えられる。

以上紹介したブートストラップ法を使って相関係数の信頼係数 95 パーセントの信頼区間の下限値 (0.025) と上限値 (0.975) を求め、付表 1 (その 1, その 2) にまとめた。付表には、誤差を少なくするために各  $\alpha$  に対してブートストラップ法を 100 回繰り返し行い、その平均、標準偏差、最小値、最大値を示した。

付表 1 (その 1) から、ステューデント化した Boot-t 法とフィッシャーの  $z$  変換した Boot-t (Fisher) 法の信頼区間の下限の推定値 (0.025) は、表 4 で与えら

れた信頼区間の理論値と比較していずれも小さい値をとり、 $\alpha$  の値が小さくなるにつれてその傾向はますます大きくなる。そして、 $\alpha = 0.1$  のとき相関係数の信頼区間の下限値は負の値をとるようになる。結局、これらの推定法は正規分布にかなり近い  $\alpha = 9, 25$  のとき以外は良い推定値を与えない。

付表 1（その 2）から、BCa 法による推定は、影響関数を使うデルタ法（influence f.）とジャックナイフ法（jackknife）が非常によく似た結果を示す。推定の精度をより詳しく調べるために、95% 信頼区間の長さ（ $= \hat{r}(0.975) - \hat{r}(0.025)$ ）と、下限値の差（=信頼区間の下限値（0.025）の平均 - 理論値）、および上限値の差（=信頼区間の上限値（0.975）の平均 - 理論値）を表 8 に与えた。表から、BCa 法は下限値の差が負の値となり、表 4 の理論値と比較していずれも小さい値を示し、また 95% 信頼区間の長さも理論値より大きくなるのがわかる。

一方、パーセンタイル法は、すべての  $\alpha$  に対して信頼区間の下限値を少し大きく推定し、信頼区間の長さを若干小さく推定する傾向が認められる。しかし、すべての  $\alpha$  の値に対して BCa 法より良い推定値を与え、5 つの推定法の中で一番良さそうに思える。

#### 4 ABC 法による信頼区間の構成と calibration による補正効果

BCa 法による信頼区間の推定は計算回数が多くなるという欠点を持つ。そこで、Diciccio and Efron [2] によって考案された、BCa 法のブートストラップを数値微分で置き換える、ABC（Approximate Bootstrap Confidence interval または Approximate BCa interval）法を紹介する。

オリジナル標本を  $x = (x_1, \dots, x_n)$  とし、推定量を  $\hat{r} = r(x)$  とする。そのブートストラップ標本  $x^* = (x_1^*, \dots, x_n^*)$  に対して、推定量  $\hat{r}^* = r(x^*)$  をリサンプリング確率  $P^*$  を使って

$$\hat{r}^* = T(P^*)$$

で表す。ここで、 $P^* = (P_1^*, \dots, P_n^*)$  は次のように定義される。

表 8 推定精度

	理論値			Boot- t 法		
$\alpha$	長さ	0.025	0.975	長さ	下限値の差	上限値の差
0.1	0.38374	0.58681	0.97055	4.5986	-4.1357	0.07915
0.5	0.24227	0.67700	0.91927	0.4751	-0.2231	0.00980
1	0.20439	0.69991	0.90430	0.2966	-0.0885	0.00373
4	0.16758	0.72163	0.88921	0.1927	-0.0244	0.00073
9	0.15883	0.72725	0.88608	0.1753	-0.0156	0.00086
25	0.15361	0.73066	0.88427	0.1644	-0.0105	0.00033

	Boot- t (Fisher)			パーセンタイル法		
$\alpha$	長さ	下限値の差	上限値の差	長さ	下限値の差	上限値の差
0.1	1.75128	-1.36930	-0.00176	0.35323	0.04705	0.01654
0.5	0.49787	-0.25792	-0.00231	0.23759	0.01220	0.00752
1	0.30016	-0.09791	-0.00215	0.20063	0.00768	0.00392
4	0.19160	-0.02548	-0.00146	0.16549	0.00264	0.00056
9	0.17291	-0.01515	-0.00107	0.15760	0.00163	0.00041
25	0.16203	-0.00975	-0.00134	0.15226	0.00122	-0.00012

	BCa (influence f.)			BCa (jackknife)		
$\alpha$	長さ	下限値の差	上限値の差	長さ	下限値の差	上限値の差
0.1	0.44334	-0.05113	0.00847	0.44623	-0.05424	0.00824
0.5	0.28023	-0.04709	-0.00914	0.28255	-0.04997	-0.00969
1	0.22404	-0.02793	-0.00828	0.22459	-0.02877	-0.00857
4	0.17369	-0.01105	-0.00494	0.17316	-0.01035	-0.00477
9	0.16237	-0.00671	-0.00316	0.16187	-0.00595	-0.00291
25	0.15560	-0.00489	-0.00289	0.15525	-0.00418	-0.00254

$$P_i^* = \frac{\#\{x_i^* = x_i\}}{n}.$$

このとき、オリジナル標本は  $P^0 = (1/n, \dots, 1/n)$  で表され、

$$\hat{r} = T(P^0)$$

である。推定量  $r(x)$  が  $X_1$  と  $X_2$  の標本相関係数のとき、リサンプリング確率

$P^* = (P_1^*, \dots, P_n^*)$  に対して

$$T(P^*) = \frac{\sum P_i^* (x_{i1} - \bar{x}_1^*) (x_{i2} - \bar{x}_2^*)}{\sqrt{\sum P_i^* (x_{i1} - \bar{x}_1^*)^2} \sqrt{\sum P_i^* (x_{i2} - \bar{x}_2^*)^2}}$$

となる。ここで、

$$\bar{x}_1^* = \sum P_i^* x_{i1}, \bar{x}_2^* = \sum P_i^* x_{i2}, x_i = (x_{i1}, x_{i2}), 1, \dots, n.$$

$T$  の 1 次微分 (影響関数) は、 $i$  th 単位ベクトル  $e_i = (0, 0, \dots, 0, 1, 0, \dots, 0)'$

を用いて、

$$\dot{T}_i = \lim_{\varepsilon \rightarrow 0} \frac{T((1-\varepsilon)P^0 + \varepsilon e_i) - T(P^0)}{\varepsilon}, \quad i = 1, \dots, n$$

によって定義する。同様に、2次微分も次のように定義する。

$$\ddot{T}_i = \lim_{\varepsilon \rightarrow 0} \frac{T((1-\varepsilon)P^0 + \varepsilon e_i) - 2T(P^0) + T((1-\varepsilon)P^0 - \varepsilon e_i)}{\varepsilon^2}.$$

なお、 $T$ の1次と2次微分の実際の計算は、 $\varepsilon$ の値を0.00001のように十分小さく定めて数値微分を行う。

このとき、ABC法による100 $a$ パーセント点は

$$\hat{r}_{ABC}[a] = T(P^0 + \lambda \frac{\dot{T}(P^0)}{\hat{\sigma}})$$

となり、信頼係数100(1-2 $a$ )パーセントの信頼区間は $(\hat{r}_{ABC}[a], \hat{r}_{ABC}[1-a])$ である。ここで、 $\hat{\sigma}$ はノンパラメトリックデルタ法の標準偏差を表し、

$$\hat{\sigma} = \sqrt{\sum_{i=1}^n \dot{T}_i^2 / n^2}$$

である。また

$$\lambda = \frac{\hat{z}_0 + z(a)}{(1 - \hat{a}(\hat{z}_0 + z(a)))^2}$$

であり、 $\hat{z}_0$ と $\hat{a}$ はBCa法で定義した偏り修正とaccelerationである。ABC法ではこれらの値を $T$ の1次と2次の微分を使って次のように数値計算する。

$$\hat{z}_0 = \hat{a} - \frac{\hat{b}}{\hat{\sigma}} + \hat{c}_q,$$

ただし、

$$\hat{a} = \frac{1}{6} \frac{\sum \dot{T}_i^3}{(\sum \dot{T}_i^2)^{3/2}}, \quad \hat{b} = \frac{\sum_{i=1}^n \ddot{T}_i}{2n^2},$$

$$\hat{c}_q = \lim_{\varepsilon \rightarrow 0} \frac{T\left((1-\varepsilon)P^0 + \frac{\varepsilon \dot{T}}{n^2 \hat{\sigma}}\right) - 2T(P^0) + T\left((1-\varepsilon)P^0 - \frac{\varepsilon \dot{T}}{n^2 \hat{\sigma}}\right)}{\varepsilon^2}.$$

ABC 法を 100 回繰り返し行い、その平均、標準偏差、最小値、最大値を付表 1 (その 2) の最後の列に記した。当然の結果であるが、ABC 法による信頼区間の推定値はBCa 法とそれほど変わらない。

ABC 法やブートストラップ法などで求めた信頼区間の近似精度を改良する方法としてcalibration というテクニックが知られている (Diciccio and Efron[3], Efron and Tibshirani[5])。以下でこのテクニックについて紹介しよう。

相関係数  $\rho$  の 100  $a$  パーセント点の推定値を  $\hat{r}[a]$  として、

$$\beta(a) = \text{Prob}\{\rho < \hat{r}[a]\}, 0 < a < 1$$

と定義する。この関数  $\beta(a)$  は calibration curve と呼ばれる。推定値が正しければ、すべての  $a(0 < a < 1)$  に対して  $\beta(a) = a$  が成立するはずである。ところが、推定値に偏りがある場合は  $\beta(a)$  を使って信頼区間を補正できる。たとえば、 $\beta(0.020) = 0.025$ , かつ  $\beta(0.980) = 0.975$  と仮定すれば、信頼係数 95% の信頼区間 ( $\hat{r}[0.025]$ ,  $\hat{r}[0.975]$ ) の補正值は ( $\hat{r}[0.020]$ ,  $\hat{r}[0.980]$ ) である。

ところが、一般に  $\beta(a)$  は未知なので、ブートストラップ法を使って次のように推定する。

$$\hat{\beta}(a) = \text{Prob.}\{r < \hat{r}[a]^*\}$$

ここで、 $r$  はオリジナル標本  $x$  の相関係数、 $\text{Prob.}$  は  $x$  のブートストラップ標本による確率、また  $\hat{r}[a]^*$  はブートストラップ標本にもとづいた相関係数の 100  $a$  パーセント点を表す。

実際の計算では、オリジナル標本から  $B$  個のブートストラップ標本を生成して、それぞれのブートストラップ標本  $x_b^*$  に対して、 $r$  に等しくなる  $a$  の値  $\hat{a}^*(b)$  を次の式から定める。

$$\hat{r}[\hat{a}^*(b)] = r$$

このとき、 $\beta(a)$  は  $\hat{a}^*(b)(b = 1, \dots, B)$  をデータとする経験分布関数を使って次のように推定できる。

$$\hat{\beta}(a) = \frac{\#\{\hat{a}^*(b) < a\}}{B}, 0 < a < 1$$



まず、ABC 法の場合に、calibration による補正が正しく行われるか調べよう。 $B = 2000$  個のブートストラップ標本を生成し、各ブートストラップ標本  $x_b^*$  から ABC 法を使って  $\hat{a}^*(b)$  を求め、 $\alpha = 25, 9, 4, 1, 0.5, 0.1$  の calibration curve を付図 2 に作成した。さらに表 9 の ABC 法の欄には、ABC 法による信頼係数 95% の信頼区間の下限と上限の推定値および calibration によるそれらの補正値を示した。なお、左端の理論値は比較しやすいように表 4 から転記した。

付図から、信頼区間の上限値 ( $a = 0.975$ ) 付近では、すべての  $\alpha$  の値に対して calibration curve と直線 ( $\beta a = a$ ) がほぼ一致して補正する必要がない。実際、表 9 の補正値の欄をみてもほとんど修正されていない。また、正規分布に近い  $\alpha = 25, 9$  の場合には、 $\hat{\beta}(a) \doteq a$  となり、すべての  $a$  に対してほとんど補正の余地がない。ところが、 $\alpha = 4$  の場合には、 $\hat{\beta}(0.014) = 0.025$  となり、信頼区間の下限値は 0.72555 から 0.71181 に補正される。すなわち、ABC 法の信頼区間の下限は理論値に比べて若干大きく推定され、calibration はこれらの値を正しい方向に修正しているにもかかわらず、修正し過ぎて理論値より少し小さくなっている。さらに  $\alpha$  の値が小さくなれば、 $a$  の値が小さいところで calibration curve と直線の乖離が大きくなる。そして、 $\alpha = 0.5, 0.1$  と小さくなるにしたがって、 $a = 0.025$  の推定値は理論値よりも小さく推定されているにもかかわらず、さらに下方に推定しなおしてますます悪い推定値を与える（表 9 参照）。このように

表 9 calibration による補正値

$\alpha$	理論値		ABC法		パーセンタイル法	
	a		推定値	補正値	推定値	補正値
0.1	0.025	0.58681	0.51758	0.19171	0.60503	
	0.975	0.97055	0.98907	0.98247	0.98680	0.98261
0.5	0.025	0.67700	0.66412	0.60403	0.70138	0.60062
	0.975	0.91927	0.91659	0.91199	0.92593	0.91396
1	0.025	0.69991	0.70060	0.66346	0.71202	0.67895
	0.975	0.90430	0.89986	0.89926	0.90668	0.90336
4	0.025	0.72163	0.72555	0.71181	0.73535	0.71641
	0.975	0.88921	0.88618	0.89032	0.89258	0.88951
9	0.025	0.72725	0.72746	0.72551	0.72631	0.70848
	0.975	0.88608	0.88619	0.88619	0.88672	0.88558
25	0.025	0.73066	0.73116	0.73026	0.73407	0.73052
	0.975	0.88427	0.88248	0.88459	0.88202	0.88149

$\alpha$  の値が小さい場合に、calibration テクニックが利用できない原因が、シミュレーションの当てはまりの悪さにあるのか、それとも正規性が大きく崩れたためなのか、はよくわからない。なお、ブートストラップの回数を  $B = 10000$  回と増やしてもこの状況は変わらない。

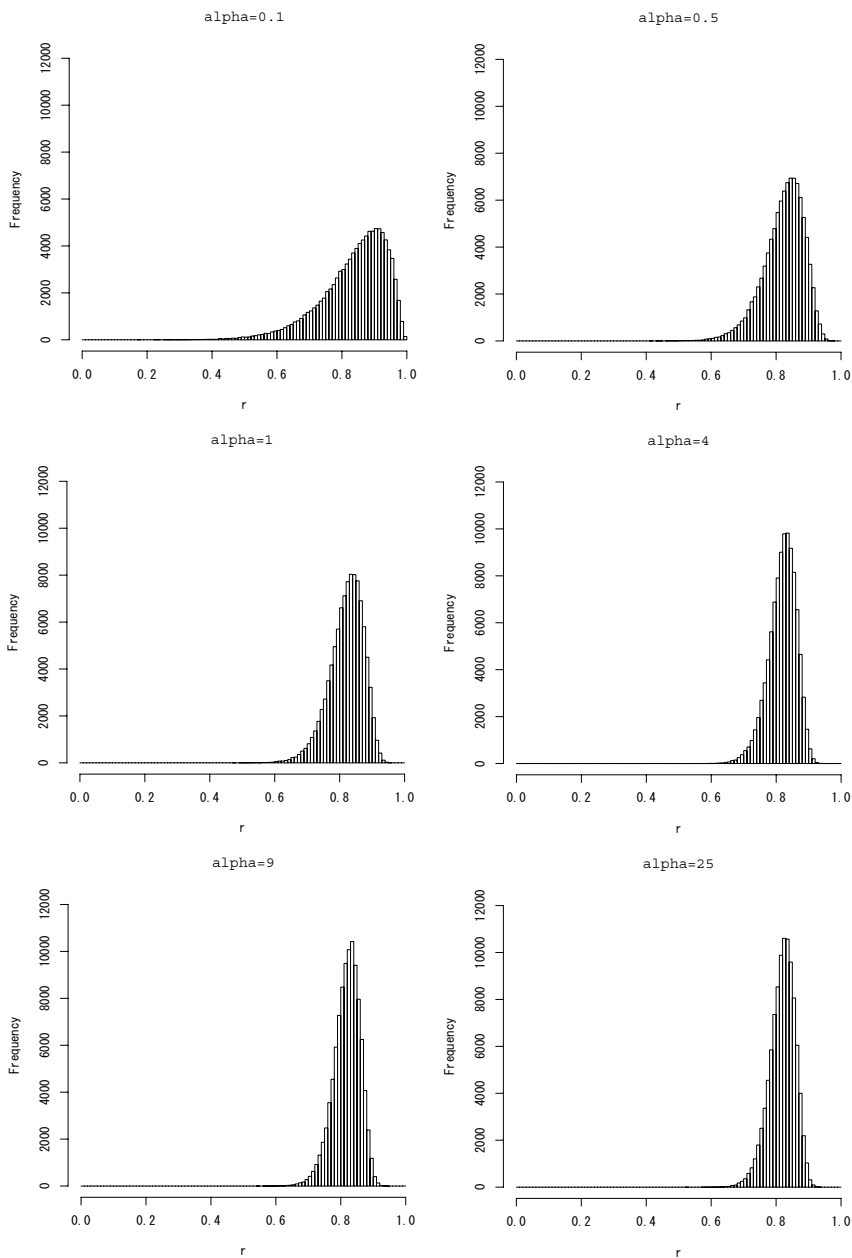
つぎに、パーセンタイル法による calibration curve を作成する。この場合、 $B = 2000$  個のブートストラップ標本  $x_b^*$  に対して、 $\hat{a}^*(b)$  を求めるためにパーセンタイル法をそれぞれ 2000 回実行すれば、合計  $2000 \times 2000$  回の膨大な数の反復計算が必要になることに注意する。 $\alpha = 25, 9, 4, 1, 0.5, 0.1$  の calibration curve を付図 3 に作成し、パーセンタイル法による  $a = 0.025$  と  $a = 0.975$  の推定値とそれらの補正値を表 9 にまとめた。付図から、すべての  $\alpha$  に対して、パーセンタイル法の calibration curve の直線からのずれは、ABC 法のずれより少し大きくなる傾向が認められるが、大体よく似た傾向を示している。 $\alpha = 25$  の場合には、 $\hat{\beta}(0.022) = 0.025$  より、信頼区間の下限 0.73407 を 0.73052 に改良できる。しかし、 $\alpha = 9$  の場合には、 $\hat{\beta}(0.0125) = 0.025$  となり、信頼区間の下限 0.72631 を 0.70848 に推定しなおし、結果的に少し悪くなっている。また、 $\alpha = 4, 1, 0.5$  の場合には、 $a = 0.025$  の値を正しい方向に補正しているが、 $\alpha$  の値が小さくなるにしたがって補正の値が大きくなり、かえって理論値より悪くなる。さらに、 $\alpha = 0.1$  のように正規性が大きく崩れると、 $a$  の値が小さいところで calibration curve が利用できなくなり、たとえブートストラップの回数を増やしても改善できないのは ABC 法の場合と全く同じである。

結局、少なくともこの論文で考えたモデルにおいては、信頼区間のブートストラップ推定値の calibration による補正は、正規性があまり崩れない場合でもそれほど有効に働かず、一方、正規性が大きく崩れる場合はまったく利用できないことがわかった。それに対して、パーセンタイル法はすべての  $\alpha$  に対して良い推定値を与え、紹介したブートストラップ法の中では一番頑健であるように思われる。

### 参考文献

- [1] Davison,A.C. and Hinkley,D.V. (1997) *Bootstrap methods and their application*. Cambridge University Press.
- [2] Diccio,T.J. and Efron,B. (1992) More accurate confidence intervals in exponential families. *Biometrika* **79**, 231-245.
- [3] Diccio,T.J. and Efron,B. (1996) Bootstrap confidence intervals. *Statistical Science* **11**, 189-228.
- [4] Efron,B. (1982) *The Jackknife, the Bootstrap, and Other Resampling Plans*. Number 38 in CBMS-NSF Regional Conference Series in Applied Mathematics. Philadelphia: SIAM.
- [5] Efron,B. and Tibshirani,R.J.(1993) *An introduction to the Bootstrap*. New York: Chapman & Hall.
- [6] Gayen,A.K. (1951) The frequency distribution of the product-moment correlation coefficient in random samples of any size drawn from non-normal universes. *Biometrika* **38**, 219-247.
- [7] 丹後 俊郎 (1983) 『医学への統計学』朝倉書店.
- [8] 松田 忠之 (2003) データ解析の基礎とデータの整理方法 日本放射線学会 近畿部会雑誌, Vol.8 No.3, 33-51.

付図1 相関係数のヒストグラム



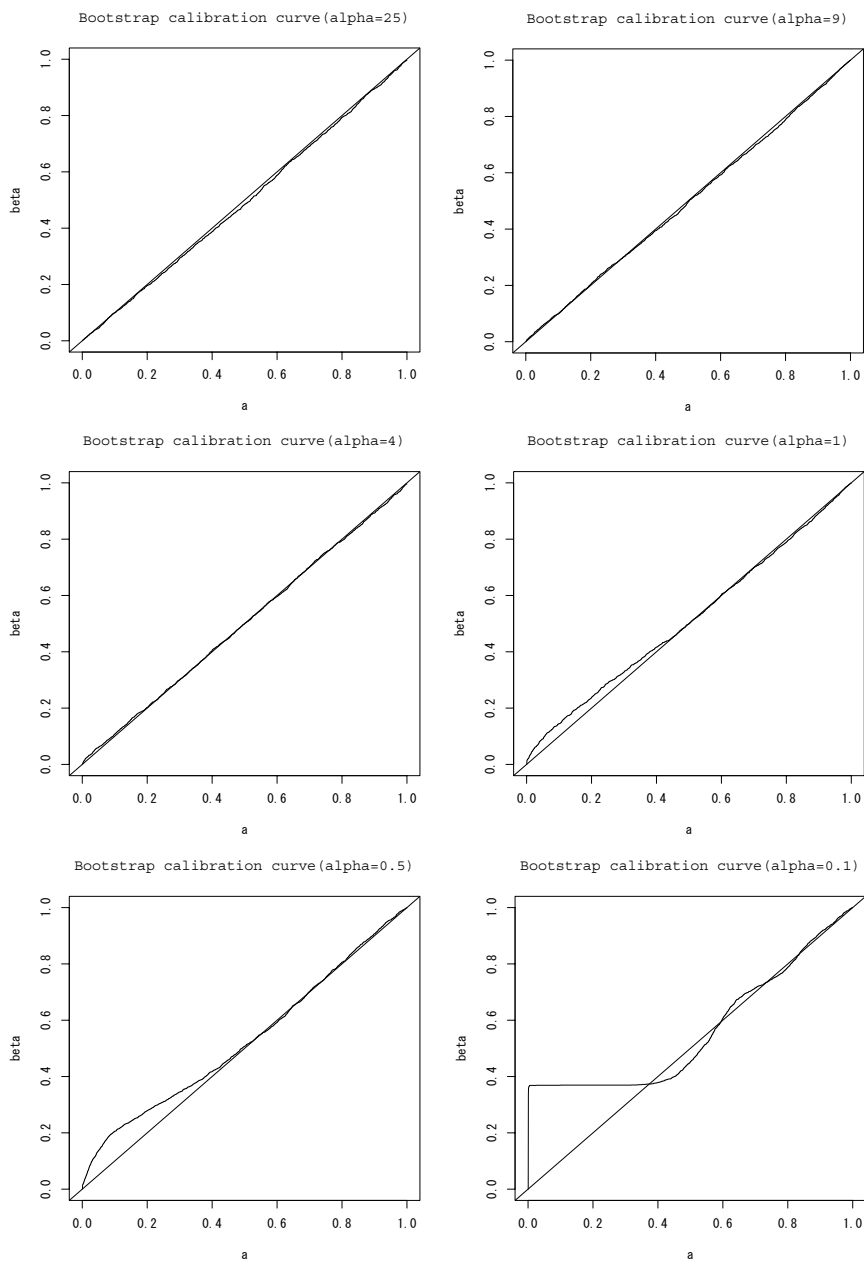
付表1 ブートストラップ推定（その1）

	統計量	boot-t 法		boot-t (Fisher)		パーセンタイル法	
$\alpha$	信頼区間	0.025	0.975	0.025	0.975	0.025	0.975
0.1	平均	-3.5489	1.0497	-0.7825	0.9688	0.6339	0.9871
	標準偏差	3.0908	0.0171	0.1650	0.0043	0.0113	0.0053
	最小値	-17.5969	1.0137	-0.9959	0.9576	0.6085	0.9697
	最大値	-0.7037	1.1064	-0.3508	0.9809	0.6606	0.9968
0.5	平均	0.4539	0.9291	0.4191	0.9170	0.6892	0.9268
	標準偏差	0.0794	0.0098	0.1021	0.0064	0.0142	0.0067
	最小値	0.2741	0.9047	0.1624	0.9007	0.6444	0.9129
	最大値	0.6057	0.9498	0.6118	0.9305	0.7245	0.9477
1	平均	0.6114	0.9080	0.6020	0.9022	0.7076	0.9082
	標準偏差	0.0383	0.0064	0.0509	0.0049	0.0146	0.0043
	最小値	0.4670	0.8922	0.3912	0.8895	0.6640	0.8970
	最大値	0.6804	0.9258	0.6822	0.9129	0.7380	0.9180
4	平均	0.6973	0.8899	0.6961	0.8877	0.7243	0.8898
	標準偏差	0.0206	0.0053	0.0235	0.0047	0.0097	0.0044
	最小値	0.6342	0.8767	0.6274	0.8758	0.7041	0.8796
	最大値	0.7380	0.9015	0.7410	0.8985	0.7455	0.9008
9	平均	0.7116	0.8869	0.7121	0.8850	0.7289	0.8865
	標準偏差	0.0158	0.0053	0.0174	0.0047	0.0099	0.0043
	最小値	0.6500	0.8723	0.6400	0.8719	0.7029	0.8771
	最大値	0.7412	0.8977	0.7436	0.8948	0.7525	0.8974
25	平均	0.7202	0.8846	0.7209	0.8829	0.7319	0.8841
	標準偏差	0.0119	0.0057	0.0121	0.0051	0.0084	0.0043
	最小値	0.6812	0.8725	0.6843	0.8716	0.7007	0.8744
	最大値	0.7489	0.8991	0.7500	0.8967	0.7479	0.8943

付表1 ブートストラップ推定 (その2)

	統計量	BCa (influence f.)		BCa (jackknife)		ABC 法	
$\alpha$	信頼区間	0.025	0.975	0.025	0.975	0.025	0.975
0.1	平均	0.5357	0.9790	0.5326	0.9788	0.5420	0.9876
	標準偏差	0.0299	0.0066	0.0309	0.0066	0.0163	0.0081
	最小値	0.4434	0.9576	0.4321	0.9573	0.5033	0.9668
	最大値	0.5925	0.9927	0.5910	0.9926	0.5853	1.0079
0.5	平均	0.6299	0.9101	0.6270	0.9096	0.6334	0.9102
	標準偏差	0.0313	0.0063	0.0308	0.0062	0.0278	0.0056
	最小値	0.5265	0.8979	0.5292	0.8970	0.5661	0.8980
	最大値	0.6825	0.9290	0.6788	0.9287	0.6883	0.9271
1	平均	0.6720	0.8960	0.6711	0.8957	0.6745	0.8965
	標準偏差	0.0300	0.0042	0.0292	0.0042	0.0260	0.0037
	最小値	0.5857	0.8837	0.5861	0.8833	0.5780	0.8851
	最大値	0.7221	0.9055	0.7212	0.9041	0.7188	0.9064
4	平均	0.7106	0.8843	0.7113	0.8844	0.7121	0.8841
	標準偏差	0.0162	0.0042	0.0157	0.0042	0.0156	0.0037
	最小値	0.6733	0.8725	0.6760	0.8728	0.6723	0.8747
	最大値	0.7414	0.8928	0.7409	0.8929	0.7424	0.8919
9	平均	0.7205	0.8829	0.7213	0.8832	0.7204	0.8829
	標準偏差	0.0136	0.0043	0.0132	0.0043	0.0132	0.0036
	最小値	0.6727	0.8730	0.6757	0.8732	0.6692	0.8736
	最大値	0.7442	0.8940	0.7443	0.8947	0.7469	0.8901
25	平均	0.7258	0.8814	0.7265	0.8817	0.7268	0.8809
	標準偏差	0.0094	0.0053	0.0092	0.0054	0.0093	0.0043
	最小値	0.6910	0.8701	0.6942	0.8702	0.6994	0.8701
	最大値	0.7458	0.8943	0.7468	0.8956	0.7507	0.8912

付図2 ABC法の場合の calibration curve



付図3 パーセンタイル法の場合の calibration curve

